



## **Statistical Analysis of NCES Datasets Employing a Complex Sample Design**

### **Module Objectives**

- Describe the major techniques for calculating estimates and their standard errors using NCES complex sample survey data
- Explain and illustrate how these techniques are applied using selected software packages
- Present and explain selected techniques for univariate and multivariate statistical analyses of NCES data using appropriate survey weights and estimation methods for calculating standard errors

### Introduction to Standard Errors

- Definition of standard errors
- Importance of standard errors
- Calculating standard errors with data from complex sample surveys

### Standard Errors

---

**What are standard errors?**

Every estimate calculated from a sample - such as a mean, a percentage, or a regression coefficient - has a standard error (SE) associated with it

---

Hypothesis testing, calculation of confidence intervals, and modeling that use complex survey data all require the calculation of SEs using appropriate methods that account for the sampling design

---

The standard deviation of the estimate's sampling distribution

---

For example, many samples could be drawn from a population to calculate a mean. The SE indicates how much the mean would fluctuate from sample to sample

---

## Importance of Standard Errors

### Why are standard errors important?

The relationship between the standard error and precision is an inverse one. A correctly computed smaller SE indicates that the estimate is more precise

An inaccurate SE can lead to incorrect findings

An underestimated SE can lead to identification of a statistically significant result where none is present (Type I Error)

An overestimated SE can lead to the failure to detect a significant finding when there is one (Type II Error)

## Calculating SEs Using Simple Random Sampling (SRS)

- An SRS gives all members of the sampling frame an equal probability of selection
- Formula for the SE for the mean of a variable  $x$  from an SRS:

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

where  $s$  is the standard deviation of  $x$  and  $n$  is the sample size

- Calculating the SE for NCES studies with complex sampling designs requires special procedures and formulas

### Calculating SEs for Datasets with Complex Sample Designs

- In complex sample survey designs, different subgroups of the sample frame members have different probabilities of being selected into the sample
- Use methods of variance estimation that take the complex survey sampling design into account
- There are two common types of methods for computing SEs for estimates from complex survey data:
  - Replication techniques
  - Taylor Series linearization

### Calculating SEs for Datasets with Complex Sample Designs (Continued)

- Examples of software that support variance estimation techniques for complex sample survey designs: SAS - SUDAAN, Stata, SPSS (Taylor Series only), R, WESVAR, and AM
- The survey-specific modules within the DLDT will provide details regarding how to appropriately calculate standard errors for your analytic purposes using the selected dataset

### **Replication Techniques for Complex Sample Variance Estimation**

- Replication techniques calculate appropriate SEs based on differences between estimates from the full sample and a series of subsamples called “replicates” that are drawn from the full survey sample
- There are two major types of replication techniques used by NCES studies: Balanced Repeated Replication (BRR) and Jackknife Replication
  - Both techniques compute replicate estimates of a statistic in different ways, as dictated by the requirements of the complex sampling design
  - They both require a set of “replicate weights” that are included in the datasets
  - NCES datasets specify the replication technique that should be implemented
- Survey-specific demonstrations of appropriate replication techniques are presented in the applicable DLDT survey-specific modules

### **Taylor Series Linearization Method**

- The Taylor Series linearization method does not require replicate weights
- It uses variables that describe the complex sample survey design, which are included in the dataset to be used to produce a linear approximation for the estimate of interest
  - Primary sampling unit (PSU) and strata
- The variance of the linear approximation is then estimated using standard variance formulas

### Examples of Standard Error Adjustment

Major techniques for calculating complex sample dataset estimates and their standard errors  
Select from the four software packages below to see how a statistic might be obtained using both SRS and complex survey methods

- [SAS](#)
- [SPSS](#)
- [STATA](#)
- [R](#)

### Critical Reminders

- Get to know your data by running descriptive statistics using weights with unweighted n's displayed in the output
- Use sampling weights, as NCES survey samples were not selected using SRS
- All statistical tests should be based on SEs that are calculated to account for the complex sample design for the particular NCES datasets you are using
- Calculate SEs using replication or Taylor Series linearization methods
- **If you do not use appropriate variance estimation techniques, you will underestimate your variance. Therefore, your tests of statistical significance, confidence intervals, and modeling results will be incorrect**

### Module Summary and Resources

#### Summary

- Described the major techniques for calculating estimates and their standard errors in NCES complex sample survey data
- Explained and illustrated how these techniques are applied using selected software packages
- Presented and explained selected techniques for univariate statistical analyses of NCES data using appropriate survey weights and estimation methods for calculating standard errors

#### Resources

- [SAS](#)
- [SPSS](#)
- [STATA](#)
- [R](#)